

基于残差网络的三维模型检索算法

李荫民^{1,2} 薛凯心^{1,2} 高赞^{1,2,3} 薛彦兵^{1,2} 徐光平^{1,2} 张桦^{1,2}

(天津理工大学计算机视觉与系统教育部重点实验室 天津 300384)¹

(天津理工大学天津市智能计算及软件新技术重点实验室 天津 300384)²

(齐鲁工业大学(山东省科学院)山东省人工智能研究院 济南 250014)³

摘要 近年来,基于视图的 3D 模型检索已经成为计算机视觉领域的重点研究方向。3D 模型检索算法包括特征提取和检索算法两个部分,且鲁棒的特征对于检索算法起着决定性的作用。目前,研究者们已经提出了许多人工设计特征和深度学习特征,但是很少有人比较它们的异同。因此,文中对不同的人工设计特征和深度学习特征的性能进行了评估分析,基于充分对比的前提,采用了多个数据集、多样的评价标准和不同的检索算法进行了实验,并进一步比较了深度网络不同层特征对性能的影响,从而提出了基于残差网络的三维模型检索算法。在多个公开数据集上的实验表明:1)残差网络所提取的深度特征相较于传统特征,综合性能提升了 1%~20%;2)与 VGG 网络所提取的深度特征相比,残差网络的综合性能提升了 1%~5%;3)VGG 网络中不同层特征的性能也有差异,深层特征与浅层特征相比,综合性能提升了 1%~6%;4)随着网络深度的增加,残差网络所提取的特征的综合性能得到了有限提高,且比其他对比特征均更加鲁棒。

关键词 3D 模型检索,特征提取,人工特征,深度特征,残差网络

中图分类号 TP391.4 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2019.03.022

3-D Model Retrieval Algorithm Based on Residual Network

LI Yin-min^{1,2} XUE Kai-xin^{1,2} GAO Zan^{1,2,3} XUE Yan-bin^{1,2} XU Guang-ping^{1,2} ZHANG Hua^{1,2}

(Key Laboratory of Computer Vision and System of Ministry of Education, Tianjin University of Technology, Tianjin 300384, China)¹

(Tianjin Key Laboratory of Intelligence Computing and Novel Software Technology, Tianjin University of Technology, Tianjin 300384, China)²

(Institute of AI, Shandong Computer Center, Qilu University of Technology (Shandong Academy of Science), Jinan 250014, China)³

Abstract In recent years, view-based 3D model retrieval has become a key research direction in the field of computer vision. The 3D model retrieval algorithm includes feature extraction and model retrieval where robust features play a decisive role in retrieval algorithm. Up to now, the traditional hand-crafted features and deep learning features were proposed, but very few people systematically compare them. Therefore, in this work, the performance of different artificial design features and deep learning features was evaluated and analyzed. Based on the premise of full comparison, multiple data sets, multiple evaluation criteria, and different search algorithms were used to conduct experiments. The effects of different layers of deep network on performance were further compared, and a 3D model retrieval algorithm based on residual network was proposed. Several conclusions could be obtained from the experimental results on multiple public datasets. 1) When comparing the deep learning features of VGG network and residual network with traditional hand-crafted features, the improvement of comprehensive performance can reaches 3% to 20%. 2) Compared with the deep features extracted by VGG network, the comprehensive performance of the residual network is increased by 1% to 5%. 3) The performance of different layer features in the VGG network is also different, and the comprehensive performance of the deep and shallow features is increased by 1% to 6%. 4) As the depth of the network increase, the overall performance of the extracted features of the residual network has limited improvement, and is more robust than other contrasting features.

到稿日期:2018-07-11 返修日期:2018-09-21 本文受国家自然科学基金(61872270,61572357),天津市应用基础与前沿技术研究计划(14JCZDJC31700),天津市自然科学基金(13JCQNJC0040)资助。

李荫民(1994—),男,硕士生,CCF 会员,主要研究方向为计算机视觉和模式识别,E-mail:18300601870@163.com;薛凯心(1993—),女,硕士生,CCF 会员,主要研究方向为计算机视觉和模式识别;高赞(1980—),男,博士,教授,博士生导师,主要研究方向为图像/视频分析、信息检索和机器学习,E-mail:gaozan114@126.com(通信作者);薛彦兵(1979—),男,硕士,副研究员,硕士生导师,主要研究方向为计算机视觉和模式识别;徐光平(1977—),男,博士,副教授,硕士生导师,主要研究方向为计算机视觉和存储编码;张桦(1962—),女,博士,教授,博士生导师,主要研究方向为视频分析、虚拟现实。

Keywords 3D model retrieval, Feature extraction, Hand-crafted features, Deep features, Residual network

1 引言

随着3D技术的快速发展、计算机硬件性能的大幅提高以及软件功能的日益强大,许多领域的3D模型都在不断增加,如游戏模型、医学模型、建筑模型等。面对现实生活中多样的3D模型,相关检索技术变得越来越重要,关于3D模型检索的研究在近年来也备受关注,其在现实世界中显示出越来越重要的实用价值。

许多早期的3D对象检索工作都是基于3D模型的,基于模型的特征描述子是由分析三维模型的物理特性(如物体形状或特定的空间几何)得到的,这种特征可以用直方图或由表面法线和曲率组成的特征袋^[1]表示。Gao^[2]提出了空间结构循环描述符,它可以保留三维模型的全局空间结构,且具有旋转及缩放不变性。Li等^[3]利用Zernike矩、傅里叶描述符、深度信息和基于Ray的特征(ZFDR)^[4]设计了混合三维形状描述符ZFDR。但是,这些基于模型的方法对模型的完整性要求很高,增加了检索难度的同时也限制了这些方法的应用范围。由于早期方法的诸多限制以及多个视图在表示3D模型时所具有的高度区分性,基于视图的3D对象检索方法成为了目前的主流检索方法。研究人员提出了一些用于图像检索的视觉特征,例如,Daras等^[5]提出了一种紧凑的多视图描述符,该描述符具有三维模型连接和全局变形的不变性,以及评估描述符(Elevation Descriptor, ED)^[6]、密集多视图描述符(Compact Multi-View Descriptors, CMVD)^[7]、视觉特征包(Bag of Visual Features, BoVF)^[8]等。相对于基于模型的结构特征,基于视图的描述符在描述3D模型方面更加鲁棒。在视图特征的基础上,许多基于视图的3D模型检索方法被提出,例如Hausdorff距离(Hausdorff Distance, HAUS)^[9]和最近邻(Nearest Neighbor, NN)^[10]方法。最近邻方法是用两个模型在不同视角下得到的视图中的最近距离来表示两个模型之间的相似度,然后将要检索的模型与目标数据库中所有物体之间的视图距离从小到大依次排序,从而得到最终的检索结果。Hausdorff距离方法和最近邻方法很相似,其不同点在于Hausdorff距离是用两个模型在所有视角下得到的视图中的最远距离来表示两个模型之间的相似度。基于概率模型的方法有Ansary^[11]所用的贝叶斯模型、自适应视图聚类方法(AVC)^[12]和自由相机视角方法(CCFV)^[13]。Ansary采用贝叶斯模型计算不同模型之间的相似性。AVC运用自适应聚类方法对3D模型的多个视图进行聚类,选出每个类最具代表性的视图,然后通过学习每个类的贝叶斯概率模型来选出代表待查询对象的最佳类,进而与数据库中的模型进行匹配。CCFV方法的提出消除了静态相机阵列对视图捕捉和三维模型检索的约束。为了进一步提升检索性能,研究人员提出了图匹配的思想,该思想被广泛应用于利用多视图信息和潜在上下文的三维模型检索中^[14-15]。Liu等^[16]提出多模型团图匹配方法(Multi-model Clique-Graph Matching, MCG),将多个视角的匹配过程转换为子群图的匹配,通过比较图节点的相似性来完成检索。Lu等^[17]提出将多视角图像转为图模型的

顶点,然后通过图匹配算法来找到视角之间的相关性。加权二分图匹配算法(Weighted Bipartite Graph Matching, WBGM)^[18]、光谱匹配(Spectral Matching, SM)^[19]和加权随机步长匹配算法(Reweighted Random Walks Matching, RRWM)^[20]都属于图匹配算法,它们通过对模型的多视角图像进行聚类,选择出代表性的视图,然后通过计算每个聚类的比例来初始化权重,并通过逐步计算视角匹配度来更新代表视角的权重,最后通过带权匹配算法完成检索。不同的是,在这些算法中,WBGM基于2D视角构建出加权二分图,SM和RRWM则利用图像中物体的轮廓和纹理等属性来实现图匹配。Liu等^[21]提出了将模型视角匹配过程转化为多模型组合的视图群匹配过程的3D模型检索算法,其很好地挖掘了视角之间的共性。

在对人工设计特征进行研究时发现,许多人工设计特征只能捕捉到一些片面的特征,例如纹理、轮廓、灰度等。目前有研究人员采用卷积神经网络来提取深度学习特征,而这种深度学习特征在计算机视觉的多个领域都取得了很好的性能,如图像分类、语音识别等;然而,在3D模型检索方面,基于深度学习特征的工作还较少。因此,本文首先分析深度学习和人工设计特征的性能差异;然后提出了基于残差网络的3D模型检索算法。本文的主要贡献包括:

1)详细对比分析了传统人工设计特征和深度特征在3D模型检索中的性能差异;

2)对VGG网络中不同层的特征在3D对象检索中的性能进行了比较分析;

3)提出了基于残差网络的3D模型检索算法,并比较了VGG网络和残差网络的性能差异,分析了不同深度的残差网络对算法性能的影响。

2 相关工作

特征在模型检索中起着非常重要的作用,下面对人工设计特征和深度学习特征分别进行简单介绍。

2.1 人工设计特征

在基于视图的3D模型检索算法中,每一个3D模型都对应一组不同视角下的图像。在这样一组视图集中,包含着丰富的模型信息,但同时要处理大量的视图,其时间复杂度非常高,而且这组视图中还含有大量的冗余信息。因此,要在一个模型的视图组中选出一小组能代表查询模型的特征视图。研究人员使用特殊的特征空间来表示这些特征视图。Zernike矩^[22]是一种常用的人工特征表示,具有完备性和正交性,在对视图的描述中具有旋转、平移和尺度不变性,但它更适合描述视图的形状特征而不适合描述纹理特征。同时,方向梯度直方图(HoG)^[23]也是被广泛应用的一种特征表示,具有良好的旋转不变性。

2.2 深度学习特征

由于不同的人工设计特征自有侧重,因此其不同数据集上的性能波动较大。为了获得更加鲁棒的特征表示,近年来,研究人员逐渐采用机器学习来自主提取图片特征^[24],目

前泛化性能较好的是深度学习。Jia 等^[25]的实验表明,使用卷积神经网络(CNN)所提取的特征适应于计算机视觉中的大多数领域,对视图中物体轮廓、纹理等特征的表达要优于传统的人工设计特征,且更加鲁棒。在 3D 模型检索中,Nie 等^[26]使用卷积神经网络来提取视图的特征,所提取的特征显著增加了 3D 模型检索的性能。随后的研究中,Liu 采用深度神经网络 VGGNet 来提取视图的特征,在 3D 模型检索中获得了远胜于传统特征的表现;但实验所使用的 VGGNet 的层数仅为 16,实验表明,层数更深且可以收敛的网络在对视图特征的表述上更加鲁棒,因此在该实验中使用了层数更深的 ResNet 来提取视图特征。

3 基于深度网络的 3D 模型检索算法

虽然深度学习特征近年来在多个计算机视觉领域都被广泛应用,但是在 3D 模型检索领域中,深度学习特征还没有被推广。因此,本文通过已有的深度网络(例如 VGG 网络和残差网络)进行特征提取,然后使用经典的检索算法对这些特征的性能进行评估。

3.1 基于 VGG 网络的 3D 模型检索算法

在实验中用于提取深度特征的 VGG 网络有 16 层(VGG-16),VGG-16 有 13 个卷积层,每个卷积层的单个卷积核的大小为 3×3 ,其步长为 1。在卷积层后面分别接入了池化层(共有 5 个),每个池化层采用的是最大池化操作,池化层的单个像素窗口是 2×2 ,其步长为 2。数据经过最后一个池化层处理后送入全连接层。这个网络有 3 个全连接层,前两层有 4096 个输出,后一层有 1000 个输出。因为 ImageNet 数据集有 1000 类,所以在最后一个全连接层进行分类,最后一层是 soft-max 层。在实验过程中取前两个全连接层的输出作为视图的两种深度学习特征,即 fc-6 层与 fc-7 层。

3.2 基于残差网络(ResNet)的 3D 模型检索算法

由 VGG 网络可知,随着网络层数的加深,网络在分类任务中的性能也在不断提高。但研究人员发现,当网络加深到一定程度时,网络性能会趋于饱和,若继续加深,则网络的性能便会开始下降,这种下降并不是由于参数过多或过拟合所引起的,而是由于当网络加深之后,网络的训练变得异常困难。与图 1(a)所示的部分 VGG 网络相比,残差网络为了解决更深层之后难以训练的问题,使用了如图 1(b)所示的残差结构,其在网络中加入跨层的 identity 连接,这样每一层实际要训练的目标函数就变成了减掉输入的残差。训练残差相较于直接训练目标函数更容易一些,且残差网络有利于梯度的反向传播,避免了梯度弥散。该实验用来提取特征的残差网络有 3 种(ResNet-50, ResNet-101, ResNet-152),其网络结构相似,仅是深度不同。ResNet-50 由 5 个卷积模块组成,第一个模块是一个使用 7×7 卷积核大小的卷积层;第二个模块共有 9 个卷积层,分为 3 组,每 3 个卷积后做一次残差操作,每组有 3 个卷积层,其卷积核大小分别为 $1 \times 1, 3 \times 3, 1 \times 1$;第三个模块共有 12 个卷积层,分为 4 组,同样为每 3 个卷积后做一次残差操作;第四个模块有 18 个卷积层;第五个模块有 9 个卷积层,每组卷积层的数目与卷积核大小同上。鉴于全连接层庞大的参数数目以及可能带来的过拟合,残差网络没

有使用全连接层,最后,用一个平均池化层来取代全连接层,并将这个平均池化层的输出作为视图特征。该实验中所使用的 3 种残差网络都分为 5 个卷积模块,不同的是每个模块的卷积层数量有所不同。

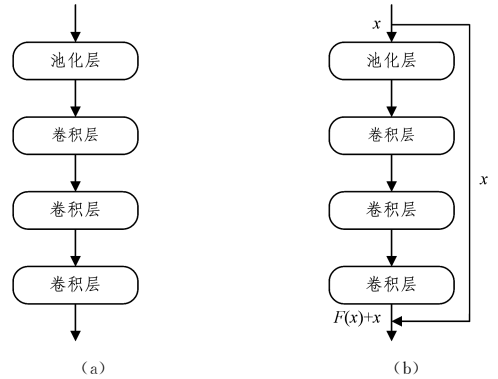


图 1 VGG 网络的部分结构和残差网络的部分结构

Fig. 1 Partial structure of VGGnet and ResNet

4 实验评估和讨论

为了充分评价基于残差网络的 3D 模型检索的性能,首先,通过不同的数据集和不同的检索算法比较了人工设计特征和深度学习特征的性能差异;然后,对深度卷积神经网络(VGGNet-16)的不同层所提取的特征的性能进行对比;其次,对两个不同深度的神经网络 ResNet-50 和 VGGNet-16 做了对比分析;最后,评估分析了不同层数的残差网络对算法性能的影响。

4.1 实验数据集

实验中使用了 3D 模型检索中广泛使用的数据集,一个是真实物体 3D 数据集,另一个是 3D 模型数据集。数据集的具体信息如图 2 所示。

(1)ETH^[27]3D 物体数据集。该数据集共有 8 个类,每个类都有 10 个物体,而每个物体都由从不同角度所拍摄的 41 张视图来表示,部分物体图像如图 2(a)所示。

(2)NTU60^[28]3D 模型数据集。该数据集分为 47 个类,每个类的物体数量不等,共有 549 个物体,每个物体由不同角度所拍摄的 60 张视图表示,部分物体的视图如图 2(b)所示。

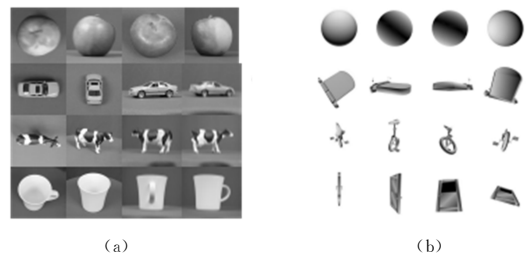


图 2 ETH 数据集的部分数据和 NTU 数据集的部分数据

Fig. 2 Partial data of ETH dataset and NTU dataset

4.2 评价标准

为了充分评估不同算法的性能,该实验采用了 7 个常用的评估标准^[29-32]来评价检索性能。

(1)查全-查准率曲线(Precision-Recall Curve, PR)。PR 曲线能够对检索性能进行综合性的评价,表示的是查全率与

查准确率的变化关系,性能的好坏通过曲线下的面积大小来进行评估,PR 曲线与两轴所围成的区域面积越大,则该算法的性能越好。

(2)最近邻评价指标(Nearest Neighbor, NN)。NN 评估结果中返回的第一个物体即检索结果中相似度最高的物体所属类别的检索精确率。

(3)第一层级正确率(First Tier, FT)。FT 表示检索结果中的前 K 个物体与待检索物体属于同一类的准确率。

(4)第二层级正确率(Second Tier, ST)。ST 表示检索结果中的前 $2K$ 个物体与待检索物体属于同一类的准确率。

(5)F 统计量(F-measure)。F-measure 是查全率与查准率的加权调和平均。

(6)折扣累积增益(Discounted Cumulative Gain, DCG)。DCG 与检索序列顺序有关,排序越靠前的检索结果的权值越大,反之则权值越小,即得到的排名越靠前的检索物体的类别正确或错误的影响力强,而靠后的影响力弱。

(7)平均归一化检索评分(Average Normalized Modified Retrieval Rank, ANMRR)。这是一种基于排序的评价方法,其值越低意味着性能越好,而其他评价指标则是越大越好。

4.3 参数设置

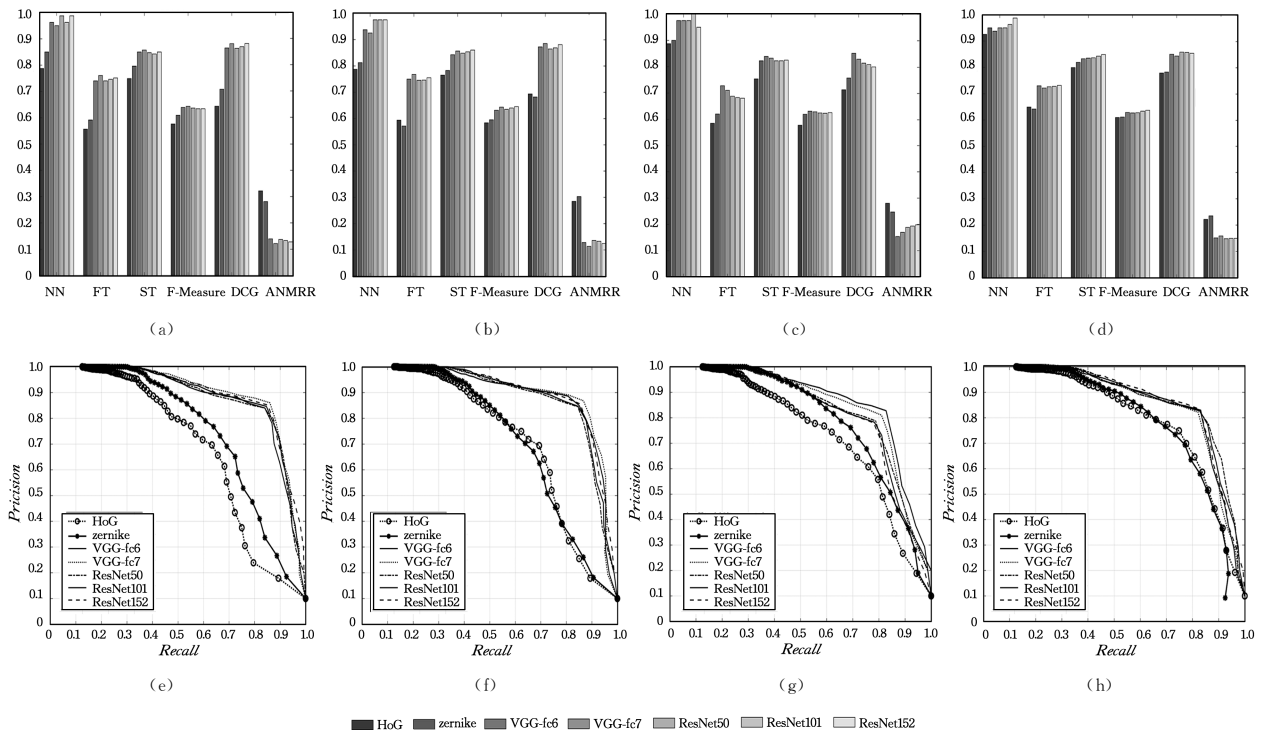
在 3D 模型检索中,关键的两步是特征提取和模型检索。在特征提取中,选取了 Zernike 矩和 HoG 特征这两种传统的人工设计特征与深度特征做对比。提取深度特征时,使用了两种在 ImageNet 大型数据集上预训练好的网络(VGG 网络和残差网络)来提取特征;在 VGGNet-16 中,提取 fc-6 和 fc-7 层的输出作为视图特征,其特征维数都为 4096;在残差网络(ResNet-50, ResNet-101, ResNet-152)上,均提取最后的 pool5 层的

输出作为视图的特征。从 pool5 层提取的特征的维数为 2048。为了充分对特征进行评估,在实验中选取了 4 种不同的检索方法进行模型检索,分别为 NN, WBGM, CCFV, HAUS, 所用检索算法的超参数使用原作者的默认设置。

4.4 深度学习特征与人工设计特征的性能比较

在该实验中,基于充分对比的前提,深度特征提取自 VGGNet-16 的 fc-7 层,所用人工设计特征是 HoG 和 zernike 特征。为了充分比较深度学习特征和人工设计特征的性能差异,使用了 4 种不同的检索算法和 2 个不同的 3D 数据集,实验结果如图 3 和图 4 所示。

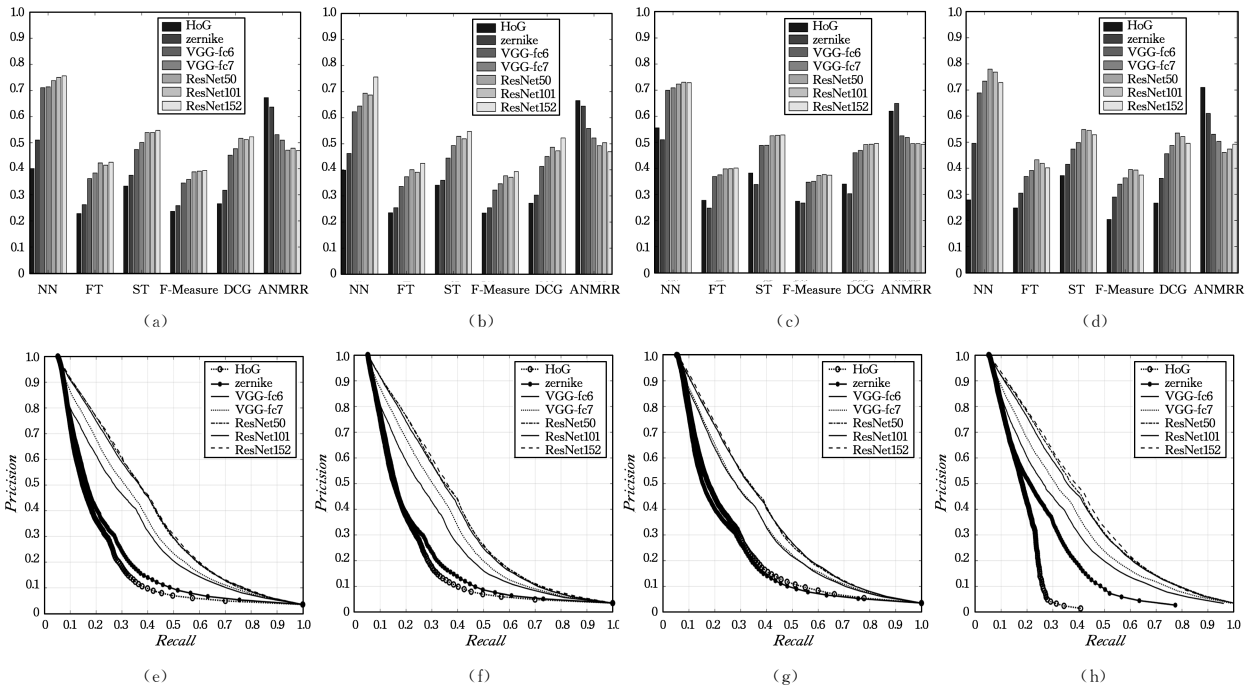
从图 3、图 4 可以看出,无论是 ETH 还是 NTU 数据集,深度特征的性能均优于人工设计特征的性能。例如,在 ETH 上,使用 WBGM 算法进行 3D 模型检索时,深度特征的性能相对于人工设计特征,在 NN, FT, ST, F-Measure, DCG, ANMRR 评估准则上分别有 10%~16%, 16%~20%, 6%~10%, 3%~7%, 17%~23%, 15%~20% 的提升;而使用 CCFV 算法时,深度特征的性能相对于人工设计特征,在 NN, FT, ST, F-Measure, DCG, ANMRR 评估准则上分别有 0~3%, 7%~8%, 1%~3%, 1%~2%, 6%~7%, 6%~8% 的提升。类似地,在 NTU 数据集上,同样使用 WBGM 算法进行检索时,深度特征的性能相对于人工设计特征,在 NN, FT, ST, F-Measure, DCG, ANMRR 评估准则上分别有 20%~31%, 12%~15%, 12%~16%, 10%~12%, 15%~21%, 12%~16% 的提升;而使用 CCFV 算法进行检索时,在 NN, FT, ST, F-Measure, DCG, ANMRR 评估标准上分别有 23%~45%, 8%~14%, 8%~13%, 7%~16%, 12%~22%, 10%~20% 的提升。



注:(a)-(d)使用了 NN, FT, ST, F-measure, DCG, ANMR 5 种评估指标,(e)-(h)是这 7 种特征在不同检索方法下的 PR 曲线

图 3 7 种不同特征和不同检索算法在 ETH 数据集上的性能

Fig. 3 Performance of seven different features and different retrieval algorithms on ETH dataset



注:(a)–(d)使用了 NN, FT, ST, F-measure, DCG, ANMRR 5 种评估指标, (e)–(h)是这 7 种特征在不同检索方法下的 PR 曲线

图 4 7 种不同特征和不同检索算法在 NTU 数据集上的性能

Fig. 4 Performance of seven different features and different retrieval algorithms on NTU dataset

最后,通过图 3 中的 PR 曲线可以看出,深度学习特征的曲线明显高于人工设计特征。相比于传统的手绘特征,深度学习特征是由神经网络自我学习、自我调整而成,这种神经网络对 3D 物体的投影、轮廓以及纹理的“理解”远胜于传统的人工特征。

4.5 VGG 网络中不同全连接层特征的性能比较

在 VGGNet-16 网络中,分别提取了 fc-6 层与 fc-7 层的特征来进行检索性能的比较,实验结果如图 3 和图 4 所示。从图 3 中可以看出,在 3D 物体数据集上,fc-7 层特征相比于 fc-6 层特征在不同数据集和不同检索方法上的表现均更佳。例如,在 WBGM 算法的实验结果中,在 NN, FT, ST, F-Measure, DCG, ANMRR 评估指标上分别有 1.2%, 2.0%, 0.7%, 0.4%, 1.6%, 1.7% 的提升;在 CCFV 算法的检索实验结果中,在 NN, FT, ST, F-Measure, DCG, ANMRR 评估指标上分别有 1.2%, 0.8%, 0.3%, 0.1%, 0.6%, 0.7% 的提升。在 NTU 数据集上,使用 WBGM 算法进行检索实验时,NN, FT, ST, F-Measure, DCG, ANMRR 评估指标分别有 0.3%, 2.1%, 2.7%, 1.3%, 2.4%, 2.1% 的提升。在 NTU 数据集上,使用 CCFV 算法进行检索实验时,NN, FT, ST, F-Measure, DCG, ANMRR 评估指标分别有 4.5%, 2.3%, 2.4%, 2.4%, 3.1%, 2.7% 的提升。此外,在图 4(e)中,fc-7 的 PR 曲线明显高于 fc-6 的 PR 曲线。在其他检索算法和其他数据集上,也可以得到相同的结论。由实验可以看出,随着层数的加深,深层网络能够提取到更加丰富的特征。

4.6 残差网络和 VGG 网络的性能比较

在该实验中,分别提取了 VGGNet-16 的 fc-7 层和 ResNet-50 的 pool5 层的特征进行比较,实验结果如图 3 和图 4 所示。从图 3 中可知,在 ETH 上,使用 WBGM 算法进行检索时,残差网络的性能比 VGG 网络的性能更优越,在 NN,

FT, ST, F-Measure, DCG, ANMRR 评价指标上分别有 3.7%, 0.1%, 0.1%, 0.2%, 0.1%, 0.2% 的提升。而在使用 CCFV 算法进行检索的实验结果中,NN, FT, ST, F-Measure, DCG, ANMRR 评估指标分别有 0.2%, 0.5%, 0.1%, 0.2%, 1.5%, 0.2% 的提升。由图 4 中可知,在 NTU 数据集上使用 WBGM 算法进行检索的实验结果中,NN, FT, ST, F-Measure, DCG, ANMRR 评估指标分别有 2.3%, 3.8%, 3.7%, 2.8%, 3.9%, 3.8% 的提升。而使用 CCFV 算法进行检索的实验结果中,NN, FT, ST, F-Measure, DCG, ANMRR 评估指标分别有 4.5%, 4.1%, 5.0%, 3.1%, 4.7%, 4.3% 的提升。此外,通过图 3 和图 4 中的 PR 曲线还可以看出,通过残差网络进行特征提取时,无论采用哪种检索算法,基于残差网络的 3D 模型检索算法的性能都是最好的。

从多个数据集和多个检索算法的实验结果可以看出,由残差网络所提取的特征明显优于 VGG 网络所提取的特征,其原因是残差网络拥有非常深的网络以及特殊的残差学习结构,且残差网络解决了网络过深导致的难以收敛以及梯度消失的问题。从该网络中提取的视图特征能捕获不同等级下更为丰富的视图特征,且随着网络深度的增加,所获得的特征更加抽象,也拥有更加丰富的语义信息。

4.7 不同深度残差网络的性能比较

在该实验中,分别取 ResNet-50, ResNet-101 和 ResNet-152 的 pool5 层特征进行比较,实验结果如图 3 和图 4 所示。可以看出,残差网络不同深度所提取的特征在该实验中的性能总体上是随着深度加深而提高的,但差异不明显,其原因是残差网络解决了网络退化的问题,但在 50 层之后,网络精度已经达到饱和。由其在预训练数据集 ImageNet 上的准确率可以看出,随着层数的加深,性能提升了 0.6%~1.7%。但是,残差网络的检索表现要优于传统人工设计特征和从

VGG-16 提取的深度特征,因此,基于残差网络所提取的特征在 3D 模型检索中更加鲁棒。

结束语 本文首先分析比较了人工设计特征与深度特征的性能差异,然后对不同网络所提取的不同深度特征的性能进行全面评估。由多个 3D 数据集的实验结果可知:1)深度特征相较于传统特征,其特征更为优越;2)对于相同网络中的不同层所提取的特征,更深层的特征的性能更好;3)在与 VGG 网络的对比中,残差网络的性能比 VGG 网络的性能更好。其原因是残差网络拥有非常深的网络以及特殊的残差学习结构,从该网络中提取的视图特征能捕获不同等级下更为丰富的视图特征。随着网络深度的加深,所获得的特征更为抽象,也拥有更加丰富的语义信息。

参 考 文 献

- [1] OHBUCHI R, FURUYA T. Accelerating bag-of-features sift algorithm for 3d model retrieval[C]// Proc. SAMT 2008 Workshop on Semantic 3D Media. 2008;23-30.
- [2] GAO Y, DAI Q, ZHANG N Y. 3D model comparison using spatial structure circular descriptor[J]. Pattern Recognition, 2010, 43(3):1142-1151.
- [3] LI B, JOHAN H. 3D model retrieval using hybrid features and class information[M]. Kluwer Academic Publishers, 2013.
- [4] VRANIC D V, SAUPE D. 3D model retrieval[D]. Leipzig: University of Leipzig, 2004.
- [5] DARAS P, AXENOPOULOS A. A 3D Shape Retrieval Framework Supporting Multimodal Queries[J]. International Journal of Computer Vision, 2010, 89(2-3):229-247.
- [6] SHIH J L, LEE C H, WANG J T. A new 3D model retrieval approach based on the elevation descriptor[J]. Pattern Recognition, 2007, 40(1):283-295.
- [7] LU K, HE N, XUE J, et al. Learning View-Model Joint Relevance for 3D Object Retrieval[J]. IEEE Transactions on Image Processing A Publication of the IEEE Signal Processing Society, 2015, 24(5):1449-1459.
- [8] OHBUCHI R, FURUYA T. Scale-weighted dense bag of visual features for 3D model retrieval from a partial view 3D model[C]// IEEE, International Conference on Computer Vision Workshops. IEEE, 2009:63-70.
- [9] DUBUISSON M P, JAIN A K. A modified Hausdorff distance for object matching[C]// IEEE Conference on Computer Vision and Image Processing. 1994:566-568.
- [10] GAO Y, DAI Q. View-based 3-D object retrieval[M]. Holland: Morgan Kaufmann, 2014.
- [11] ANSARY T F, DAOUDI M, VANDEBORRE J P. A Bayesian 3-D Search Engine Using Adaptive Views Clustering[J]. IEEE Transactions on Multimedia, 2006, 9(1):78-88.
- [12] ANSARY T F, DAOUDI M, VANDEBORRE J P. A Bayesian 3-D Search Engine Using Adaptive Views Clustering[J]. IEEE Transactions on Multimedia, 2006, 9(1):78-88.
- [13] GAO Y, TANG J, HONG R, et al. Camera Constraint-Free View-Based 3-D Object Retrieval[J]. IEEE Transactions on Image Processing, 2012, 21(4):2269-2281.
- [14] LIU X, WANG M, YIN B C, et al. Event-Based Media Enrichment Using an Adaptive Probabilistic Hypergraph Model. [J]. IEEE Transactions on Cybernetics, 2015, 45(11):2461.
- [15] ZHU L, SHEN J, JIN H, et al. Content-Based Visual Landmark Search via Multimodal Hypergraph Learning[J]. IEEE Transactions on Cybernetics, 2015, 45(12):2756.
- [16] LIU A A, NIE W Z, GAO Y, et al. Multi-Modal Clique-Graph Matching for View-Based 3D Model Retrieval[J]. IEEE Transactions on Image Processing, 2016, 25(5):2103-2116.
- [17] LU K, HE N, XUE J, et al. Learning View-Model Joint Relevance for 3D Object Retrieval[J]. IEEE Transactions on Image Processing A Publication of the IEEE Signal Processing Society, 2015, 24(5):1449-1459.
- [18] GAO Y, DAI Q, WANG M, et al. 3D model retrieval using weighted bipartite graph matching[J]. Signal Processing Image Communication, 2011, 26(1):39-47.
- [19] LEORDEANU M, HEBERT M. A Spectral Technique for Correspondence Problems Using Pairwise Constraints[C]// Tenth IEEE International Conference on Computer Vision. IEEE Computer Society, 2005:1482-1489.
- [20] CHO M, LEE J, LEE K M. Reweighted random walks for graph matching[C]// European Conference on Computer Vision. Springer, Berlin, Heidelberg, 2010:492-505.
- [21] LIU A, WANG Z, NIE W, et al. Graph-based characteristic view set extraction and matching for 3D model retrieval[J]. Information Sciences, 2015, 320:429-442.
- [22] KHOTANZAD A, HONG Y H. Invariant Image Recognition by Zernike Moments[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 1990, 12(5):489-497.
- [23] DALAL N, TRIGGS B. Histograms of oriented gradients for human detection[C]// IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2005(CVPR 2005). IEEE, 2005:886-893.
- [24] PERRONNIN F, MENSINK T. Improving the fisher kernel for large-scale image classification[C]// European Conference on Computer Vision. Springer-Verlag, 2010:143-156.
- [25] DONAHUE J, JIA Y, VINYALS O, et al. DeCAF: A Deep Convolutional Activation Feature for Generic Visual Recognition [C]// International Conference on Machine Learning. 2014:647-655.
- [26] NIE W, CAO Q, LIU A, et al. Convolutional deep learning for 3D object retrieval[J]. Multimedia Systems, 2017, 23(3):325-332.
- [27] LEIBE B, SCHIELE B. Analyzing appearance and contour based methods for object categorization[C]// 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. IEEE, 2003, 2:II-409.
- [28] CHEN D Y. On visual similarity based 3d model retrieval[C]// Computer Graphics Forum. 2003:223-232.
- [29] GAO Y, WANG M, JI R, et al. 3-D Object Retrieval With Hausdorff Distance Learning [J]. IEEE Transactions on Industrial Electronics, 2013, 61(4):2088-2098.
- [30] GAO Y, WANG M, ZHA Z J, et al. Less is More: Efficient 3-D Object Retrieval With Query View Selection[J]. IEEE Transactions on Multimedia, 2011, 13(5):1007-1018.
- [31] LLER H, LLER W, SQUIRE D M, et al. Performance evaluation in content-based image retrieval: overview and proposals [J]. Pattern Recognition Letters, 2001, 22(5):593-601.
- [32] NIE W Z, LIU A A, SU Y T. 3D object retrieval based on sparse coding in weak supervision[J]. Journal of Visual Communication & Image Representation, 2016, 37(C):40-45.